

Questionnaires in translation

Harkness, Janet; Schoua-Glusberg, Alicia

Postprint / Postprint

Sammelwerksbeitrag / collection article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Harkness, J., & Schoua-Glusberg, A. (1998). Questionnaires in translation. In J. Harkness (Ed.), *Cross-cultural survey equivalence* (pp. 87-126). Mannheim: Zentrum für Umfragen, Methoden und Analysen -ZUMA-. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-49733-1>

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Questionnaires in Translation

JANET A. HARKNESS AND ALICIA SCHOUA-GLUSBERG

Translation of questionnaires is the most frequently chosen route to implementing 'equivalent' instruments in cross-national and cross-lingual survey research. The article presents the framework of current survey translation practice: the various procedures proposed for translation and for assessment of translation products and the respective advantages or disadvantages of each. In doing so, pointers are made to research gaps in questionnaire adaptation and evaluation for cross-cultural work and to the need for interdisciplinary transfer from cognitive survey research, translation studies and statistical analysis in order to establish a thorough-going methodology of questionnaire adaptation, assessment and documentation.

1. Why and when questionnaires are translated

The most common reason for translating questionnaires is to be able to field an instrument not available in the language required for fielding. Thus the best-known cross-national survey projects operating on a regular basis (including EUROBAROMETER, ISSP, WVS, LATINOBAROMETER)¹ translate from source questionnaires into the other languages required. Within countries with more than one official language, questionnaires for the different linguistic populations are usually produced from one questionnaire. In America, Spanish-speaking populations are frequently interviewed using Spanish questionnaires translated from the English source questionnaires.

¹ EUROBAROMETER: the official regular survey carried out for the European Commission; ISSP: the International Social Survey Programme, an annual survey of topics of interest to the social sciences, 30 member countries in 1998; WVS: World Values Survey; LATINOBAROMETER: the South American counterpart to the EUROBAROMETER.

The need to translate a questionnaire is sometimes apparent from the outset if one or more targeted populations is known to need a different language from the one in which the questionnaire is/will be designed. Alternatively, the need for a translation may only become apparent at a later stage. In some American studies, for example, the ‘luck of the (sample) draw’, i.e., where the sample falls, decides whether a translation is made. A multi-stage probability sample which selects rural counties at the first stage may, for example, end up including counties with a high density of Spanish speakers who require a Spanish questionnaire. Translations are also made in some surveys on an *ad hoc* basis during fielding; interviewers orally translate their questionnaire in order to field with respondents who require another language (see section 4.7).

Instruments are also translated when researchers wish to field items originally conducted in another language. Two further reasons are a) when questionnaires are translated so as to consider their items or coverage in developing new questionnaires and b) when translations (or glosses) of items are made, usually into English, for the electronic question banks and databases now appearing. The issues related to these last two contexts are not discussed here.

2. Materials Used to Produce Translations

2.1 Source Language Questionnaires (SLQs)

A common point of departure for translation is what we call a *source language questionnaire* (SLQ) in finalised form. In a *finalised* questionnaire, every component has basically been decided and fixed. In European multi-national and in international projects, the SLQ is often in English and is finalised before translation begins. One notable exception is the EUROBAROMETER, for which French and English source questionnaires are provided. Occasionally, translation begins when the SLQ is still at the drafting stage. The aim here may be to use *advance translating* (section 4.5) to refine the draft towards a final version.

In some studies, there may not *be* a questionnaire to translate. Instead, topics, dimensions, and perhaps numbers of items may be set out in one language; the questionnaire is then developed in another language on the basis of these. Although elements of ‘translation’ of concepts are involved in this situation (cf. Gutknecht and Rölle, 1996:297f.), it is best thought of as *foreign language implementation of design specifications*. In this situation, a questionnaire in the language of the specifications may never appear, or only appear at a later stage to allow designers to discuss the implementation.

2.1.1 Development of SLQs

SLQs for multi-lingual implementation are developed under different conditions, which in turn may impact on the products. They include the following:

- mono-culturally by people all sharing the same general cultural-linguistic background;
- by people from one country with different first languages or habitually using different languages (Switzerland, Canada);
- by people from one country with different standard varieties of one language (in the UK, the Welsh, Scots, N. Irish and English);
- by people from several countries speaking varieties of one language (Spanish-speaking South American countries, or GB, USA, and New Zealand);
- by people from different nations and cultures speaking different languages using one language as a lingua franca (e.g., the ISSP, multi-lingual, multi-cultural);
- (potentially related to the above) by a group developing an SLQ in a language which is the first language of very few or no-one in the group.

The processes and dynamics of multi-cultural development raise numerous issues, not least since questionnaire development involves detailed consideration of formulations. These cannot be gone into here.

2.1.2 Types of Finalised SLQs

Finalised SLQs take various forms:

- a questionnaire exactly in the format fielded in a country using the source language. It may indeed be some country's questionnaire;
- a questionnaire *text* in the source language which is not set out as a questionnaire and also not pre-tested in the source language;

The wording for this text is thus fixed but format and lay-out, for example, may not be. In view of the communicative nature of all the elements in a questionnaire (Harkness 1995, 1996; Schwarz, 1996), this may not be an optimal source document from which to produce a translated questionnaire expected to be 'equivalent'.

- an SLQ which includes background variable items to be used everywhere or, as a further variation, an outline of required background variable information (cf. Harkness, Mohler and McCabe (1997) on background variables and cross-cultural comparability);
- an *annotated* questionnaire;

Recent ISSP modules have annotated questionnaires.² The questionnaire now distributed in English for a module is the 'prototype' questionnaire for the Programme. It cannot be used *exactly as it stands* in any country. Countries fielding in English would, for example, remove certain notes in brackets and insert their country-specific elements in the same way as countries required to translate. The ISSP annotation includes:

- notes on elements to be adapted in a country-specific manner, e.g., school-leaving ages and culture-specific institutions such as *Parliament*;
- general (non-country-specific) glosses of elements thought or known to be problematic in translation (e.g., British English terms such as *civil servants*, *social security*, (cf. Role of Government, 1996);
- notes on dimensions in items expected to need free translation (cf. Work Orientations, 1997);
- reminders to implementers to observe certain goals of the design (cf. Religion, 1998);
- indications where an ISSP member has permission either *not* to field or to adapt a question (cf. Religion, 1998 for Japan);

- indications which, if any, questions are optional;
 - specifications of special background demographics beyond the compulsory ISSP set.
- a questionnaire which incorporates items which are to be translated and fielded in each language and/or culture (etic items) and items which are to be individually developed as cultural equivalents for all or some of the languages and/or cultures (emic items);

Despite the appeal of the increased and more detailed coverage of local information through emic items, concerns about how to establish and code 'equivalence' mean that these options have been little used in multi-national survey work (but see, for example, Przeworski and Teune, 1970; Flaherty et al., 1988; Triandis, 1994; and Johnson, this volume). Hui and Triandis (1985:143-144) outline the 'combined etic-emic approach' in which etic constructs are identified and then measured in emic ways. They, too, note that the lack of item equivalence and scalar equivalence make "direct comparison of cultures impossible".

- a 'new' SLQ which includes items already used in other studies.

Items already used in accredited studies will generally be preferred over newly developed items, partly because actual use is seen as the best possible 'pre-test' and because replicating them offers some opportunity for comparison of findings. The conditions under which such 'old' items were developed and the existence of translations which have already been used affect the questionnaire currently being translated. Although the new questionnaire may have been developed multi-culturally, 'borrowed' items may not have been. Tension then arises between tinkering with item wording to 'improve' them and using 'tried and tested' items. Existing translations, even if considered sub-optimal, may be adopted for similar reasons.

2.1.3 Draft Source Questionnaires

Different types and stages of questionnaire text are variously referred to as *draft questionnaires*. A source language draft questionnaire used in a decentering approach

² Source questionnaires are available from the ISSP secretariat and ISSP archive (addresses at the Programme's web site: www.issp.org)

(Werner and Campbell, 1970) is the text from which both source and target language versions are produced more or less simultaneously (section 4.1). In contrast, draft translations in committee translation approaches are the texts from which the final translated version emerges (section 4.3). ISSP drafting involves a series of draft questionnaires in English, produced by a multi-cultural drafting group in working towards a final source language questionnaire. Each draft incorporates feedback in English from (potentially) all member countries of the Programme.

3. Facts of Translation

Translation of instruments is not the only means available to gather information on dimensions and constructs across cultures, but it is generally seen as the only means to ensure item equivalence and scalar equivalence (cf. Hui and Triandis, 1985; Flaherty et al., 1988; Van de Vijver, this volume). Acquadro et al. (1996:575) identify two major arguments for using the same (translated) questionnaire in different countries: a) “a common international interpretation and analysis of the results is only possible if the data come from the same instrument” and b) all new data acquired about an instrument contribute to the validation and reputation of the instrument (especially relevant in the context of much-used instruments). Translation is at all events the most frequently adopted approach and certainly the approach the majority of researchers see as the most viable option (cf. Guillemin et al., 1993; Van de Vijver, this volume).

Translation beyond the field of instrument translation takes many forms, with different outputs for different purposes. The goals of a particular translation – whether this be to convey the factual information, the sound effects, or the communicative intention of a source text – determine the product of the translation process (cf. Kiraly, 1995:55). In survey research, questionnaire translations are generally, if vaguely, required to ask the same questions and offer the same response options as a source text. Rightly or wrongly, they are expected to do so by means of a close rendering of the source questionnaire (section 4.4).

3.1 Equivalence and Adequacy/Appropriateness

Languages are not isomorphic and so translation cannot be expected to operate on a one-to-one basis across languages. This means that what goes in (the source language text) cannot be completely matched by what comes out (the target language text). Indeed, a mechanistic notion of input and output is itself misleading. Moreover, translation is not solely concerned with translating ‘meaning’ (on ‘meaning’, see section 3.2). As mentioned, some translations are aimed at conveying sound effects or emotional effects, while others focus on conveying factual information or (distinct from this) communicative intention.

Translation necessarily involves difference as well as similarity. Absolute absence of difference would amount to replication of the source text in the source language, absolute absence of similarity would force us to query the status of one text as a translation of the other. In terms of ‘equivalence’ between texts, difference is sometimes discussed in terms of ‘loss’ and ‘gain’ (Newmark, 1988:7f.). *Semantic* loss and gain occurs as soon as nuances of meaning associated with a lexeme sense (see Lyons (1977:197f.) on sense) in the source language are not covered by the lexeme(s) in the target language, while other nuances, anchored in the target language and culture, are ‘gained’. An example of *grammatical* loss or gain could be that the sex of individual people referred to is indicated in one language (as frequently is the case in German) but not in another language. Harkness (1996a) discusses issues this raises for translating survey items, while Acquadro et al. (1996:582) recommend gendered versions of questionnaires.

Social science research has its own rich array of kinds of ‘equivalence’ (Johnson, this volume). Different but equally varied kinds of equivalence are referred to in translatology writings (e.g., Snell-Hornby, 1988:13-22; Kiraly, 1995:54f.). These include expressions which formally match some used in survey research but have different senses (e.g., ‘functional equivalence’). Gutt (1991:10-17) argues against applying the term equivalence to translations at all and demonstrates that equivalence (however understood) cannot be automatically equated with quality (see, too, Reiss and Vermeer, 1984; Hönig and Kussmaul, 1984). We avoid using the term ‘translatory equivalence’ here. In

considering translation quality, we prefer to think of assessment in terms of appropriateness or adequacy for a given task. The appropriateness or adequacy of a given translation is then defined in terms of the degree to which it successfully fulfils stipulated goals for the translation, within the constraints of what is possible. Admittedly, a major problem here for survey research is that concrete translation goals are rarely articulated (sections 3.4 and 3.5).

3.2 Meaning is defined in and by use

The second constraint on our expectations about translation and equivalence has to do with survey standpoints on the meaning of well-written items. Handbooks outline how to avoid writing poorly formulated and ambiguous items. The implication is that the meaning of well-written items will be clear or unambiguous. This, in turn, implies that there is such a thing as ‘the’ meaning of an item.

However, in many fields of research this is by no means the currently accepted view of meaning as related to words and larger units of language in use. Over the past three decades, in research fields concerned with meaning – such as linguistics, literary theory, social psychology, language philosophy and translation studies – the meaning of words and the larger units they constitute has come to be seen as determined in and by use. Meaning in a given context is thus seen as determined in and by that context in its widest sense and as co-constructed between users. By this is meant that the co-text (the surrounding or accompanying text), the immediate and larger contexts, the text producer(s) and the recipient(s), as well as the lexical content of expressions and the propositional content, all affect what may be perceived as ‘the’ meaning of a communication in a given instance. Moreover, a reading (‘meaning’) perceived by an addressee need not be the meaning intended (and perceived) by the speaker/writer. Seeing meaning as *dialogic* shifts the goal of communicators away from making ‘the’ meaning clear and towards making the *intended* meaning clear, or as clear as possible. This applies equally to communication in questionnaires (Harkness, 1996b; Schwarz, 1996). The success of everyday communication shows that we are adept (ultimately) at getting

intended messages across. But the multiple repairs, repetitions, explanations and expansions we engage in simultaneously underline that meaning is, indeed, dialogic.

This is relevant for survey translation in two main respects. First, given that meaning is not fixed and finite, one of the goals of translation must be to convey the intended and most salient reading of a well-written question. The intended meaning of an item should therefore be documented for translators in the source materials they receive for their task (Hambleton, 1993). Whether this reading can be conveyed by means of close translation (of the moderate kind) is a separate issue (see 4.4), as is whether the salient reading in translation continues to tap the dimension or construct required. The factors which determine a given reading may differ across cultures, thus a close translation in terms of lexical content can conflict with the goal of conveying intended meaning. This poses a real problem for survey translation, since in many instances we currently do not have documentation on intended readings, nor on intended dimensions. We also lack detailed guidelines and examples for what might constitute an acceptable degree of freedom in producing the target text. In this situation, researchers understandably hesitate to experiment.

Significantly, the number of researchers and research bodies suggesting and calling for improved and systematic documentation and guidelines is growing (Hambleton, 1993 and 1994, Prieto, 1992; Guillemin et al., 1993; Acquadro et al., 1996; Van de Vijver and Hambleton, 1996). Only through such documentation and provision of information will modifications to current practice be able to be realised in a consistent fashion. Basic empirical research is needed, too, on how issues of adaptation can best be tackled beyond the modest beginnings of Harkness (1996c), Harkness et al. (1997), Mohler, Smith and Harkness (this volume) and Harkness, Mohler and McCabe (1997).

3.3 The Dual Nature of Questionnaires: Instruments and Texts

A third factor shaping the demands on questionnaire translation is the dual character of questionnaires as *texts* “destined for discourse” (Harkness, 1994, 1995) and as

instruments of measurement. Whether self-administered or read out by interviewers, in principle at least, the questionnaire text determines what is said to or read by respondents. In the closed question format, the questionnaire also basically determines the responses open to respondents. Even if rarely talked about in quite these terms, one of the goals of questionnaire design in the monolingual context is thus to optimise communication of intended stimulus and response. However, optimising communication in the target language may, again, run counter to close translation expectations. For the present, we lack research on how questionnaires as holistic entities (Harkness 1995, 1996b) can best be adapted for other language implementation.

3.4 Translators need Information and Task Specifications

Given the fluidity of meaning and the range of interpretation many texts allow, translators decide what they want (or have) to communicate and then try to do that (Wilss, 1996; Kussmaul, 1986), within the confines of what is possible across the given languages and cultures. These decisions are never made in a vacuum. If not provided with task specifications, translators are forced consciously or unconsciously to provide their own, as is evident from think aloud protocols of survey translators at work (Harkness, 1996c). Translation manuals thus increasingly stress the importance of adequate information and *task specifications* for translators (cf. Wilss, 1996; Kussmaul, 1995; Holz-Mänttari, 1984; Gutknecht and Rölle, 1996; Gile, 1995). However, task specifications for questionnaires based on requirements, guidelines and standards agreed by the survey research community are not yet available. The recommendations, overviews, and guidelines which have appeared in other areas of research using instruments (Hambleton, 1993, 1994; Prieto, 1992; Acquadro et al., 1996) and, of course, within the field of translation studies (see above) are invaluable starting points for the social sciences.

3.5 Providing Information and Task Specifications

Given the complex dual nature of questionnaires – seemingly simple texts with overt and covert measurement properties – task specifications need to be negotiated between those best informed about textual properties and those best informed about measurement

properties. These task specifications are likely to consist of a compromise between what researchers wish to have and what translation (not translators) can deliver. There can be little doubt that specifying translation tasks will require an exchange of information between researchers, questionnaire designers, target language implementers and translators. 'Rules' of practice in certain fields (Acquadro et al., 1996) suggest that that personal contact between item writers, research teams and translators is assumed to be possible. In our experience, however, implementing situations constantly arise in which a) no individual exchange will be possible, b) people involved in the design are in any case no longer sure what items 'mean' in detail, c) item writers (if, indeed, items are products of individual composition) quickly become anonymous (Harkness, 1994) and d) items move in undocumented journeys from survey to survey, country to country, and formulation to formulation.

Be this as it may, documentation could be organised without undue difficulty to provide the information needed to negotiate a translation. This would need to include information on what is required in terms of measurement, what is intended in terms of textual communication, what is possible in terms of translation versus other forms of adaptation, and where particular language and/or culture problems may arise. Certainly, this kind of documentation is essential to further interdisciplinary understanding of the demands on questionnaires in translation.

4. Some Survey Translation Procedures

In this section we briefly describe the translation approaches most frequently referred to in the survey context. Back translation is discussed under section 5 on *assessment* and not here under *translation*, although it is sometimes referred to as a 'translation method' (Sechrest et al., 1972; Brislin, 1970; McKay et al., 1996) and we briefly outline why.

Back translation involves the translation of a text which itself is a translation *back* into the original language (5.3.1). It is most commonly used and recommended as a way to assess

translation work. (e.g., Werner and Campbell, 1970; Brislin, 1970, 1976, 1980, 1986) but other uses are also suggested. Werner and Campbell (1970) describe a form of decentering which includes back translation steps and assessments – the multistage iterative process. They also suggest back translation can be used for translator assessment. Theoretically, there are as many ways to approach a translation *out* of what was originally a target language *back into* the source language as there are to produce a target language translation in the first place. Descriptions of back translation describe what the (back) translation product can be used for rather than the translatory goals and method involved in producing the back translation text itself. It is not an approach for arriving at a translation in the way that committee (parallel) translation or decentering can be seen to be (see below). We find it helpful, therefore, to maintain a distinction between kinds of translation approaches and uses to which a translation can be put.

4.1 Decentering

Decentering in translation (Werner and Campbell, 1970) is a technique which begins from a *draft* questionnaire in the source language in order to produce final questionnaires in *two* languages (source and target) through a process of paraphrase and translation between source language and target language. Paraphrase is seen as a way of decentering the text in both languages, that is, producing texts which are not ‘centred on’ or ‘anchored to’ a specific culture and language.³ Schoua (1985) reports positively on a Spanish-English decentering experiment, as do McKay et al. (1996); recent psychological test translation work has also shown interest in decentering (Tanzer et al., 1997).

Werner and Campbell (1970) suggest several approaches to decentering including taxonomic decentering, multiple stage translation, mapping of paraphrases across languages, and interview schedule-based decentering. In essence, decentering involves the following (with variations depending on the procedures chosen):

³ The idea that this is possible (in natural-sounding utterances at least) runs counter to theories in which meaning is determined by use and use is invariably tied to the culture in which it occurs.

- each draft question is reformulated and paraphrased with the goal of eliminating culture-specific aspects and simplifying complex sentences into basic, most simple constructions;
- each item (or set of paraphrases for an item) is translated into the target language. Here the idea is not to translate in any 'close' or literal fashion, but to produce as many paraphrases in the target language of the 'meaning' of the source language text(s) as possible;
- these paraphrases in the target language are translated in comparable 'paraphrase fashion' into the source language;
- the sets of paraphrases for each item/sentence in each language are compared;
- the closest equivalents across the two languages are selected;
- this selection forms the basis of both final questionnaire texts for the item/sentence.

One generally important feature of decentering approaches is that they stand in direct contrast to the 'close' translation described in section 4.4.1, which clings to words or structures across languages and, in doing so, produces unnatural-sounding translation. However, through decentering, the items may also end up sounding odd, an aspect Werner and Campbell (1970:411) consider unimportant. Another important feature of decentering is the centrality it gives to working out different versions in different languages before a 'source' text is fixed for posterity (cf. 4.5).

At the same time, in a world of survey fielding of old 'tried and tested' items, the source text is often not open to emendation. Translation may also be required into many languages. Werner and Campbell focus on two language instrument development and it is difficult to see how a many-to-many matching across, say, twenty languages might be practicable (cf. Werner and Campbell, 1970:406). In addition, the procedures are demanding in terms of time, personnel, qualifications and funding, all real stumbling blocks in the world of survey management and funding. The inherently subjective basis of judgements taken at each of the comparative steps – from identifying paraphrases to be rejected to selecting 'the best' or closest equivalents – is a key factor and, for some, a key

weakness in decentering. Lastly, decentering takes a sentence-based view of meaning as its starting point, with words and grammar interacting to provide sentences with their meaning (Werner and Campbell, 1970:401). More investigation will be needed to assess how successful the procedure can be in different contexts (perhaps in a leaner version) and how it can cater for different notions of textual equivalence, the dialogic view of meaning, and idiomatic-sounding items.

4.2 ‘Direct’ or ‘one-for-one’ translation

In terms of procedure, the ‘simplest’ and cheapest translation approach has one translator producing one translation in a traditional manner – the translator simply produces a translation to the best of her/his ability. Sechrest, Fay and Zaidi (1972) call this ‘direct translation’, a term not to be thought of as in contrast to ‘indirect’ or ‘less straightforward’. References to this kind of approach specify neither the translation process nor the product type envisaged. Limiting the work to one person is attractive in terms of funding, organisation and streamlining of time schedules. The absence of support materials for translators, the low impact so far of translatology findings on theory, practice, and on assessment procedures, the disadvantages of relying on one person’s perceptions and skills, the lack of coverage of regional differences (where these are an issue), and, finally, the data quality risks this involves are drawbacks to this approach, at least as frequently implemented (Sechrest, Fay and Zaidi, 1972; Guillemin et al., 1993; Acquadro et al., 1996).

McKay et al. (1996) use the term ‘direct translation’ for translation from source to target language, that is, ‘one way’ (forward) translation as opposed to ‘two way’ (forward and backward or ‘double’) translation, ie., translation and back translation.

Acquadro et al. (1996:577-578) define direct translation as translation which “comprises borrowings, calques (loan translations) and word-for-word translation”. (What is meant by word-for-word translation, which is contrasted with ‘literal translation’ is uncertain.) They contrast direct translation with indirect translation. This last is characterised as

involving “transposition, modulation, equivalence and adaptation”. The processes outlined for indirect translation suggest, broadly speaking, that it pursues (stipulated) goals as a covert translation, whereas direct translations are (among other things) overt translations. Covert translations read ‘naturally’, overt translations signal that they are translations (see 4.4.1).

4.3 Committee and Modified Committee Translation

Committee approaches are used for translation (discussed here) and for translation assessment (discussed in section 5.3.2). Committee or parallel translation involves several translators who make independent translations of the same questionnaire (Brislin, 1980; Schoua-Glusberg, 1992; Acquadro et al., 1996 (team translation); Guillemin et al., 1993). At a reconciliation (consensus, revision) meeting, translators and a translation co-ordinator compare the translations, reconcile discrepancies and agree on a final version which taps the best of the independent translations or, alternatively, appears in the course of discussion. The committee members should provide competence in whatever varieties of the target language are required for respondents (McKay et al., 1996; Acquadro et al., 1996) and in the various skills required for survey work (Van de Vijver and Hambleton, 1996; Johnson et al., 1997).

The committee approach is fairly labour, time and cost intensive. Schoua-Glusberg (1992) proposes a modified committee approach which involves group work but not parallel translation. Each translator works on a different part of the text rather than the whole text. The committee reviews the text provided in sections by different people and arrives at a final version. The approach can maintain the quality of parallel translation work while cutting some costs and reducing the time needed to arrive at a final version, in particular if the questionnaire is long (Schoua-Glusberg, 1992). Care must be taken to ensure that consistency is maintained across the translation. This applies, of course, for any translation, whether produced in parallel fashion, using the modified approach, or produced by one translator. However, the modified committee approach may require

particular care. For example, it is conceivable that two translators both translate the same expression in the individual parts of the questionnaire each has, and that each comes up with a suitable, but different, translation. Neither translation gives cause for discussion in the committee session. Without consistency checks as a standard part of the process, the term used to refer to 'X' in one part of the questionnaire could unintentionally end up being different from the term used for 'X' elsewhere.

Like any approach which assesses equivalence or appropriateness on the basis of textual evaluation, committee decisions are ultimately based on subjective judgements. Committees are as open to group dynamic drawbacks as other groups. Given individual competence within the group, however, group screening is likely to be effective. While competent translators are necessary (section 6), the role and skills of the committee co-ordinator are crucial, as is an understanding and acceptance of the procedures by all involved.

Institutes or researchers faced with sporadic cross-lingual implementations may find it complicated to maintain a translator committee group who stay 'in practice'. The German ISSP questionnaire is, essentially, the only translated questionnaire the institute involved produces per year. In this context, no group of skilled translators working frequently together on survey translation is available. Instead, the co-ordinator recruits institute researchers with the necessary understanding of survey instruments and grasp of English, student research assistants with competence in English, with and without survey knowledge, and two translators (skilled practitioners) – one external and working regularly as a translator, one internal, working in survey translation research. Germanic language members of the ISSP who need to translate (Austria, Germany, Norway, Sweden) also confer on problems and solutions. In this way, a mix of input from much the same group of people can be maintained from year to year without unreasonable costs or effort. This compromise solution has proved useful, in that it brings together translation drafts guided by instrument knowledge, translations from skilled practitioners and 'fresh' insights from 'outsiders' (students) and includes a degree of consultation across countries and languages.

4.4 Close and Literal Translation

Survey research often favours close renderings of questions as a means to arrive at equivalent measurement. In view of a) the often vague nature of discussions of what ‘close’ translation is and b) this vagueness notwithstanding, the differing descriptions of ‘close’ and ‘literal’ found in instrument literature (and, differently again, in translation literature), we indicate briefly the minimum kinds of ‘closeness’ we understand to be involved. A close rendering in survey terms would, for example, be expected to refer to the same entities (have the same *referent*) as referred to in the source text (sport, education, TV-watching, God). The entities would also be referred to using lexemes which cover as much of the same *sense(s)* as possible and come as close as possible lexically to the source text choice. The morning star and the evening star – if unlikely candidates for an item – may help us make distinctions here (cf. Lyons, 1977: 197f.). Thus, if the source text mentions the *morning star* (referent Venus), the target text would, if possible, refer to that too, and not, for example, to the *evening star* (referent Venus), nor to Venus with a lexeme like *Venus*. Exceptions to close renderings are what in survey research are sometimes called ‘country-specific renderings’ for country-specific institutions such *Parliament*, *A-levels* or *Prime Minister* (and, presumably, across cultures and religions, *God*). The idea that this might also apply to *sport*, *education*, and *TV-watching* is not familiar to survey researchers. If these were felt to require ‘country-specific renderings’, new items would probably be looked for which avoided these problems. Furthermore, the propositional content of the source text would be expected to be maintained in the target text. In other words, *God created human beings* (X predicate ^(create) Z) in English would not, for instance, become something more like (Z predicate ^(create) X) in another language.

4.4.1 Too Close for Comfort?

For survey translations, greatest emphasis is usually placed on avoiding differences in semantic information (lexeme senses) and grammatical information (e.g., number,

tensing, mood). Sticking close to source language (and culture) *ideas* and *concepts* in items is on occasion tricky enough, even with well-designed items. Sticking close to ideas by means of sticking close to lexical senses will at times amount to a lost cause. But even where a close rendering is possible, a 'successful' translation in terms of a close rendering carries with it no guarantee that communicative functions (and with them measurement properties) are equally well retained (Hulin, 1987; Hambleton, 1993; Flaherty, 1988; Johnson, this volume; Van de Vijver and Leung, 1997; Van de Vijver, this volume). In some contexts the need for culture-specific equivalents is apparent, in others rather less so (Harkness and Braun, in preparation).

Survey translations frequently go beyond only trying to convey ideas and concepts from the source text. Following what seems to be a survey understanding of 'close' translation, formulations, words and syntax are copied or imitated across languages (cf. what Acquadro et al. (1996) call 'direct' translation, something like a word-for-word gloss and the 'literal' translation described in McKay et al., (1996). This partly stems from the survey concern to ask the same items in order to compare data. It also reflects survey perceptions of the options available through translation. It may also be related to using back translation as an assessment (Hulin, 1987). An extreme form of close or literal translation is unlikely to result in a covert translation, that is, one which does not signal its foreign origins. It also stands in conflict with the fact that translation involves and requires change, adaptation and compromise.

Covert translation versus overt translation (House, 1977) raises questions related to how respondents perceive the questions and questionnaire. *Overt translation* is the production of a target language text which signals (in a variety of possible ways) that it is a translation. *Covert translation*, in contrast, produces a target language text which reads like an original text of the given text type in the target language and thus does not signal that it is a translation. A considerable body of cognitive research in the monocultural context documents that respondents react to features of questionnaire design which researchers have neglected, and that they do so in predictable ways (reviewed in Schwarz, 1996). By extension, we could expect that questionnaires which signal they *are*

translations (or simply come across as odd texts in some way) will prompt certain responses in respondents. Unless there is a valid reason why respondents should consider the origins of the questionnaire, we suggest that survey translations should be covert translations, ie., should read/sound like original target language questionnaires (cf. Sechrest et al., 1972; Hulin, 1987; Harkness, 1996a). But even if a close rendering of an item results in a) a natural-sounding translation which to all intents and purposes b) fulfils measurement requirements and c) is viewed as a close translation of the source item, *difference*, that is, non-equivalence, is unavoidable.

4.5 Advance Translation

Drafting procedures recommended to the ISSP (Harkness, 1995b) propose that modules are translated while still in the drafting process, before the source questionnaire is finalised. Experience has shown that many translation problems linked to source text formulations only become apparent, even to experienced cross-cultural researchers, if a translation is attempted. As necessary, source formulations can be adapted or annotated on the basis of advance translation feedback and notes for the (annotated) source document can be greatly enriched. This is often particularly relevant for the languages and cultures furthest removed from the models underlying the source text; these are otherwise unlikely to receive much consideration in notes. Nevertheless, without empirical demonstration of the need to translate in advance, the additional effort and costs involved mean it is unlikely to be adopted as a standard practice.

4.6 Passing on the Translation to Fielding Institutes

This is less an approach to translation than a way of dealing with translation as an issue. Research groups sometimes commission the fielding organisation to produce the translated questionnaires required and may or may not be involved in any of the ensuing steps of production and assessment. Fielding institutes may well have more experience in producing different language versions of questionnaires than researchers. Ultimately, however, someone decides on task specifications, guidelines and assessment procedures.

In view of the generally scant exchange of information and research findings on translation procedures and assessments, the requirements, procedures and procedure control measures should be carefully negotiated with the institutes.

4.7 Translation of Finalised Questionnaires ‘on the fly’

Translation is sometimes left up to the interviewer or an intermediary. By translating the available questionnaire orally, they are thus able to field with respondents requiring a different language, not an infrequent problem in multi-lingual societies. In the American context, if only a small number of respondents are expected to need a specific language version (not enough to ‘warrant’ producing a written translation), it is not uncommon for translations to be done ‘on the fly’, as it is called. Beyond knowing that these translations are made orally, little can be said about the *approach* taken in a specific case (e.g., free or close translation, emphasis on communicative functions, covert or overt translation, etc.).

Some modes of administration make it less likely, in the Western context at least, that translations will be done on the fly. If properly administered, a self-completion format should preclude this. Telephone interviews are more open to translation on the fly, whether as part of the design or not. Under pressure to display good interview achievement rates, interviewers may opt to translate rather than forgo an interview. Importantly, they may also have management permission to do so. Some US research companies use (readily available) bilingual telephone company operators to ‘assist’ interviewers with respondents unable or unwilling to answer in the language of the questionnaire. *Ad hoc* translation is, of course, used in other countries and continents, too. The general appeal is clear if we consider the obstacles for interviewers fielding, say, in parts of Africa or Asia, loaded down with eight and more language versions of questionnaires, but with never exactly the right version to hand.

The absence of written translations is of import for the data obtained. The relevance of standardly requiring not only a written translation of *question content* but also a *finished questionnaire* in translation is directly related to standard practices and requirements of

monolingual studies. In oral interviewing (of whatever mode) interviewers are trained to avoid providing non-standardised input in the dialogue. Despite problems this raises (Houtkoop-Steenstra, 1995; Cate Schaeffer, 1995; Stanley 1995a, 1995b), recorded questionnaires (paper or computer applications), with integrated, formulated instructions for interviewers, enable interviewers to comply with standard practices. In addition, the hard or soft copy questionnaires thus available are reasonable indications for later reference of what respondents were actually asked and offered as answer options, at least linguistically speaking. None of this follows from translation on the fly. In the worst case, researchers relinquish control of fielding and end up with response data but no record of what was asked and answered in general and in particular.

5. Assessment

The two central issues in translation assessment are what is to be assessed and how this is to be assessed. If the goals to be met by the product are not specified in advance, the criteria of assessment also cannot be specified in any manner fair to translators. In questionnaire translation, they are rarely specified, i.e., articulated, at all. Translation task specification is both a prerequisite for objective assessment of translations and for replication and validation of any assessment made. Without specifications, the usefulness of assessment procedures cannot be evaluated either. Given proper task specifications for translation, forms of assessment can be tailored to fit, within the confines of what can reasonably be expected.

Assessment of translated questionnaires is sometimes tied to the translation procedure adopted and/or to the questionnaire design (e.g., whether old items and old translations are replicated or not). Decentering has translation assessment as integral to producing the final questionnaire in *two languages*. Committee translation has assessment (reconciliation) as a central process in producing the final version of one *translated* questionnaire. Assessment may also be independent of both. Assessment procedures used once translation has been carried out are considered in section 5.3. Assessments of instrument equivalencies are discussed elsewhere (e.g., Hulin, 1987; Hulin, Dragow and

Komocar 1982; Van de Vijver and Leung, 1997; Van de Vijver, this volume; Saris, this volume).

5.1 Bilingual and Monolingual Feedback

Assessment will either be made by bilinguals, monolinguals or both. Findings from bilinguals are not automatic pointers for findings from monolinguals. The two groups perceive texts, language and cultures differently (cf. Hulin, 1987). This said, bilingual appraisal of translations is an inevitable component of translation productions (with each translator appraising as she/he formulates) and a frequent and useful component of translation product assessment. It is important to avoid pressure on assessors and translators to defend one or the other translation version (issues of criticism of colleagues, superiors, etc.). Independent bilingual assessment of a text may simply mean that people not involved in the translation assess whether they consider the translated text to be 'equivalent' to the source text. Without stipulating what equivalence is understood to involve, this is clearly a hazardous undertaking. Given the subjective nature of textual assessment, even when guidelines are provided, it is important both to ensure a spread of qualified views and to include monolingual feedback.

Monolingual judgements of a translated text should in our view only be made on texts in the language the assessor speaks. In other words, we see little to be gained from having monolinguals compare a source text and a back translation to decide on a text they cannot read procedures. Having monolinguals go through a questionnaire – either as part of a pre-test, a probe interview or simply as copy-editing readers (e.g., interviewers reviewing it for readability) – are very useful. Given that they only know the target text, this group will only be able to comment on things they are asked about or which happen to strike them. This is perhaps less systematic than the comparison which can be made by (suitable) bilinguals accustomed to assessing texts intensively. There are also limits to what can be expected of people who match the target population in terms of textual assessment. Moreover, many questions remain open as to the representativity of the information received.

In general terms, the number of people needed to gain reliable information is a problem. For procedures described in section 5.3.4, large numbers of respondents are required before the data can be considered representative. In the target setting there will presumably be enough monolinguals available. For bilinguals, wherever the testing is being done, this may not be the case. Thus finding enough candidates of suitable language proficiency to participate is one issue. Apart from the double cultural perspective of bilinguals mentioned above, bilinguals may well not match demographic characteristics of the monolingual target groups. In addition, testing along the lines of split ballots, probe interviews, simulated interviews, etc., (section 5.3.4) are all expensive, time-consuming procedures which only produce data which must then be evaluated.

5.2 Assessment Basics

Very little of 'cookbook' nature can be passed on here either about tackling survey translation or assessing survey translation quality. The social sciences have been slower in articulating needs and guidelines than has been the case in psychology and other clinical fields such as medicine or specialised research fields (Hambleton, 1993, 1994; Guillemin et al., 1993; Acquadro et al., 1996; Prieto, 1992; Van de Vijver and Hambleton, 1996). Even in these fields, however, guidelines are still fairly general on translation techniques and assessment. They also seem to imply a greater homogeneity of items and of origins of items and a greater intensity of use and re-use than is common in surveys. Moreover, little research is available on comparative assessment of translation assessment procedures themselves. The following recommendations are therefore necessarily of basic nature:

- In assessing (and in finalising) the translation avoid loss-of-face confrontations. Set up different dynamics from the start to allow open assessment and criticism;
- Assess translated questionnaires (TQs) as covert translations, that is, as texts which read/sound like questionnaires designed in the target language.

- Base TQ assessment on bilingual assessment of SLQs and TQs, defining beforehand which equivalencies are essential.
- Base TQ assessment on monolingual (target population) assessment. (This may be hampered by the same problems on lay person feedback as experienced in monocultural research).
- Keep assessment requirements realistic. A covert translation required to maintain communicative equivalence and measurement equivalence may need to be a rather free translation.
- Choose assessors who understand the mediums involved - questionnaires as instruments and as texts in translation. As need be, find the competencies in several people.
- Even if the assessment is made by one person, extend revision decision-making to a group (which should include translators).
- Budget for assessment and revision (time, people, money).
- Keep in mind that translation assessment is not an assessment of measurement reliability and validity and take steps to assess these.

Research is needed on evaluating assessment procedures. Findings from the last decade of cognitive psychology research on survey design and from translation studies are likely to be valuable here, as is recent research on translation issues in the social sciences, medicine and psychology (Wilss, 1995; Kussmaul, 1995; Dollerup and Lindegaard, 1994; Acquadro et al., 1996; Prieto, 1992; Guillemin et al., 1993; Van de Vijver and Hambleton, 1996). Work is also needed on assessing approaches to survey translation (Sinaiko and Brislin, 1973; Schoua, 1985; McKay et al., 1996; Harkness and Braun, in preparation; Harkness, 1996c) and on investigating survey translation quality (Brislin, 1970; Schoua-Glusberg, 1988; Harkness and Braun, in preparation). Finally, since translation can only deal with some aspects of instrument adaptation, translation procedures and translation assessment have to be coupled with statistical investigations of instrument measurement properties and comparability across versions. Here, too, we need to clarify how best to implement all these in sequence or iteratively.

5.3 Assessment Procedures

General types of assessment which have some currency in discussions of survey translation include *back translation*, *comprehension assessment*, and various kinds of assessments based on *analysis* of response data.

5.3.1 *Back Translation*⁴

The term back translation is used in survey research literature and in translation studies to refer to the *translation of a translation* back into the source language. Almost without exception in survey work, the purpose of back translation is to compare/contrast the back translation with the source text, usually with a view to assessing the quality of a translation.⁵ For survey translation, back translation is seen as offering a solution to the fact that researchers often need information about the quality of translations without being able to read and evaluate these themselves. It operates on the premise that if the translation is good, 'what went in ought to come out', the central idea being that a translation *back* into a language which can be understood allows researchers insight into a text in a language which cannot be understood. The basic steps involved are as follows:

- A source text in one language (Source Language Text One, SLT1) is translated into another language (Target Language Text, TLT).
- The TLT is translated back into the language of SLT1 by a second translator, unfamiliar with the SLT1 and uninformed that there is an SLT1. This second translation, the back translation, is SLT2.

⁴ This section draws on material from Harkness (1996c).

⁵ There are several references to back translation in translatology. Here authors generally include bilingual perspectives in discussion of the texts. Baker (1992) uses the term to refer to the (natural-sounding) English glosses she gives for texts in 'exotic' languages which themselves have been translated out of English. The purpose is to demonstrate difficulties met with in translations and ways of dealing with these. Vinay and Darbelnet (1958) discuss back translation as a means of assessing accuracy using bilingual insight. They suggest that back translation should really be into some third 'neutral' language, something unlikely to appeal to monolingual survey researchers looking for insight. Uses of back translation suggested in Werner and Campbell (1970) are discussed in section 4. Brislin (e.g., 1970, 1976, 1980, 1986) is most often cited in connection with back translation.

- SLT1 is compared to SLT2.
- On the basis of differences or similarities between SLT1 and SLT2, conclusions are drawn about the equivalence of TLT to SLT1.

The more identical SLT1 and SLT2 are, the greater the equivalence between the TLT and the SLT1 is considered to be. For example, if the source questionnaire in English has *Please enter your nationality* and the back translation in English has *Please enter your nationality*, then the TLT is assumed to say the same, only (somehow) in a foreign language. The frequent references in demographics literature to distinctions and overlaps between *citizenship*, *nationality*, *ethnic membership* and *religion* (e.g., Maier, 1991; Harkness, Mohler and McCabe, 1997) make clear, however, just how fluid the survey overlap for these concepts is. A point in case is the 1997 discussion in Russia about the absence of the fifth rubric in new passports (the rubric for 'nationality') and the positive reaction Russian Jews reportedly had to this. Among the admittedly scant references to translation assessment in study reports in recent years, back translation appears frequently (overview in Guillemin et al., 1993; Acquadro et al., 1996).

In general, back translation can be likened to a primitive metal detector; it can be expected to miss much, but also to pick up some things. It cannot identify what it picks up and neither, unfortunately, can the monolingual researcher. There is no necessary connection between what is 'picked up' (by virtue of being different from the SLT1) and what needs to be picked up. Pragmatically, it is likely but not necessary that major differences between a source text and its translation will also be reflected in a back translation. In saying this, note that the interpretation of what is 'major' is left as much to our readers as it is in the survey context to those deciding whether to change a translation. Brislin (1976) states, moreover, that one of the main disadvantages of back translation is that a good back translator will resolve problems actually present in the TLT (cf. Kussmaul's (1995) recommendations to translators to improve the text). Be that as it may, deciding on the presence or absence of 'difference' raises issues of meaning, appropriateness and equivalence and of how decisions are made on what constitutes a 'major' difference or a 'salient' difference.

A number of general points are noted here by way of clarification:

- Back translation itself does not deal with what, if anything, should be changed in a translation nor, crucially, how to change anything. Monolingual researchers thus come no further than looking at two texts in one language. In order to revise the target language questionnaire, bilingual competence has to be re-introduced, with all the imponderables this involves.
- If the back translation is simply used to make a list of things for *bilinguals* to look at in the target language questionnaire, other procedures which compare the SLQ and the TLT, such as committee discussions (sections 5.3.2 and 4.3), are more efficient. Since bilinguals *are* needed, the notion of monolingual and ‘objective’ insight often associated with back translation is misplaced (cf. Acquadro et al., 1996).
- At the same time, the goal of providing researchers unable to read the TLT with as much (relatively unfiltered) information as possible on the text is an important one. Research on think aloud survey translation protocols (Harkness, 1996c) suggests that these provide useful information in this situation. Even if researchers do have competence in the target language, they will be likely to welcome additional input of the kind think alouds can offer.
- Finally, we note that researchers using approaches which involve back translation *as one step* frequently describe the entire procedure as back translation. This, we suggest, indicates that researchers recognise more is needed than a back translation. Different approaches to actually producing the (back) translation seem to be involved, ranging from morpheme for morpheme, ‘literal’-and-stilted, to quick and free paraphrase (Acquadro et al., 1996; McKay et al., 1996; Werner and Campbell, 1970; Schoua, 1985).

5.3.2 Committee Assessments

Even when the translation has been produced by one translator, committee assessment is recommended (Guillemin et al., 1993; Acquadro et al., 1996; McKay et al., 1996). McKay et al. (1996), for example, describe a variety of group assessments in

experimental contexts, noting the usefulness at different stages of assessment of appraisal by monolinguals from the target population, bilinguals, survey design experts, interviewers, as well as people more narrowly seen as having the necessary translatory expertise to appraise target and source texts. Time, personnel and funds available for translation assessment are usually more restricted than in the McKay et al. experimental setting. A spread of expertise is clearly desirable, the question is what is most effective and viable. This depends on what is to be assessed. If, for example, the questionnaire is required to be a covert translation, understandable to a broad public, and has to follow, say, house-style question formats, it may be better to alternate discussion between those with translatory and survey design expertise on the one hand with feedback from people held to represent the target population. Assessment of instrument equivalence beyond translation adequacy also needs to be incorporated, at least if the intention is to modify the questionnaire on the basis of findings from statistical analysis.

5.3.3 *Comprehension Assessment*

Comprehension assessments of translations are based on the idea that if people are able to explain, describe or perform accurately on the basis of having read translated material, then the translated material accurately contains the information necessary to perform these tasks. The focus of assessment is thus on the factual information retained in a translation rather than on other aspects of equivalence or translation adequacy. These forms of assessment have been used, for example, to assess translations of instruction materials. In school 'text comprehension' testing, related procedures assess not the texts, but the *recipients* of the texts. This highlights an intrinsic source of potential error when assessment of textual adequacy is based on performance, that of discrepancies between human performance and perceptions on the one hand, and text content on the other. Brislin (1976) outlines further limitations connected to knowledge-testing and performance-testing.

In the survey context, de-briefing sessions with respondents have been used to probe their comprehension of specific items or formulations, as have focus groups in the developmental and translation stages. One advantage of these assessments is that they can be made with monolinguals. Limiting factors are the need to construct tests and questions, the costs involved, the potential impact of social desirability (and knowledge) factors, uncertainties about the representativity of input made, and the limits on the detail which can be pursued due to time, fatigue, or respondent suitability. Beyond this, too, these assessments may provide little information on ‘fine tuning’ aspects of text formulations. (Acquadro et al., 1996; McKay et al., 1996; Schoua-Glusberg, 1988, 1989).

5.3.4 Statistical Analyses

Statistical analyses take various forms and have different goals, as papers in this volume demonstrate. They investigate aspects of comparability and equivalence inaccessible through assessment of translation quality. Ultimately, what is needed is an approach which neither neglects evaluation of textual and communicative equivalence nor statistical assessment of measurement properties. As mentioned earlier, guidelines are needed on how best to combine these. Statistical analyses of item, battery, construct, or instrument equivalence use data from pre-tests or main study fielding. They investigate instrument quality from various perspectives on the basis of data produced across versions of the questionnaire. Similar distributions or response patterns are taken as evidence of either equivalence between SLQ and translation or as indicative of *instrument* equivalence, validity, reliability, etc. (e.g., Hulin, Drasgow and Komocar 1982; Hulin, 1987; Hazashi, Suyuki and Sasaki, 1992; Davis, 1993; Van de Vijver and Leung, 1997; Van de Vijver, this volume; Saris, this volume). Analyses of unexpected main study results can lead researchers to examine translations as a source of difference (Braun and Scott, this volume), or, indeed, visual representations (Smith, 1995). Facet theory analysis (Borg, this volume; Brislin, 1980) is seen a way of identifying information related to measurement which could help translators.

Procedures used to test translations include the following:

Split ballot assessments One group of bilinguals is administered the SLQ, another comparable group receives the TLQ. If the responses across the two groups have similar distributions or patternings – either marginals or more complex distributions – the questionnaires are considered to function as equivalent instruments. Alternatively, one group completes one half of the questionnaire in translation and the other half untranslated. The other group completes the other half of each questionnaire (source and target) and responses across the groups and the questionnaires are compared (Hulin, 1987; Hayashi, Suyuki and Sasaki, 1992; Acquadro et al., 1996).

Double administration tests Bilingual respondents complete the questionnaire in the SLQ and the translated version. Here, again, discrepancies across their responses are taken as indications of differences in the two versions. The remarks made earlier about differences between monolingual and bilingual responses to texts and the problem of assessing text on the basis of performance apply here, too. Moreover, what follows from finding ‘differences’ or ‘similarities’ across questionnaires remains open. Presumably, either statistical differences lead to textual examinations and these re-open the imponderables of textual assessment, or the versions are left and the data is adjusted. Double administration tests involve asking people to do something again. However, repetition itself affects responses, as research in the monolingual context has shown. Respondents asked the same questions (or who think they are asked the same questions) try to make sense of the repetition by finding new interpretations for the questions (reviews in Schwarz, 1996). It is quite possible that if asked the same things in two languages, respondents either decide that something different must be meant or decide something is behind being asked ‘the same thing’ twice. Either way, this may lead to different responses. Differences (and similarities) may thus not be related to features of the translation.

Post hoc analyses which examine translations on the basis of unexpected response distributions across languages are usually intended to help guide interpretation of results rather than the development or assessment of translation. Both the approach and the findings raise new questions about expected versus unexpected results and about

translation differences versus culturally differentiated responses (Braun and Scott, this volume).

6. Organising, Translation and Assessment

Decisions on which translation procedure to adopt and how to assess the translation are influenced by the time, funding, expertise and personnel available, as well as by specific aspects of a given study. Each of these factors impacts on the others. Planning for translation should be made early in the design stage. If translation is known to be a possible (but not certain) factor, contingency plans for this should cover details of people, payment and time schedules for translation and for assessment.

Time Organisation: The time allocation must include time for translation (including ‘time off’ before revision), assessment, revision, pre-testing, production of the final version of the translated questionnaire. If the SLQ and the translated questionnaire(s) are to be fielded simultaneously, the SLQ must be available early enough to allow for the steps above. In actuality, this is seldom the case and quality, documentation, learning curves and satisfaction suffer.

Funding Allocation: Translators in all fields of work are often poorly thought of and poorly paid. Even if translators are well-paid, translation costs are likely to be low in comparison to other costs in a survey, while poor instrument adaptation can be costly in terms of data quality. Proper selection of translators, appropriate briefing, provision of suitable materials, and adequate assessment to identify problems will contribute significantly to the success of translation products.

People and skills: Survey literature variously advocates that translators should be ‘bilinguals’, ‘professional translators’, people with knowledge of empirical social science

research, or combinations of all of these, without much indication of what, concretely, is required in terms of performance. Thus different research groups, while using similar terms, may be referring to different kinds of expertise and knowledge.

Bilingualism, for example, is a term applied to various kinds and degrees of abilities in two languages. One distinction made is between *compound* bilinguals, who learn one language after the other, and *co-ordinate* bilinguals, who learn both more or less simultaneously (cf. Wilss, 1996:206f.). Another distinguishes between bilinguals who learn a language when young and others who learn it when adult. Competencies differ in each case. Moreover, neither the degree of bilingual competence needed for survey translation nor what other competencies are needed has been empirically investigated. The high level of proficiency often glossed as 'first language proficiency' in the target language and 'good proficiency' in the source text language certainly seem to be sound requirements. The problem remains, nevertheless what is meant here by 'high level' and 'good' proficiency and how this can be assessed *before* the work is commissioned. It is important to remember, however, that some kind of 'word perfect' performance in the two languages is neither a necessary *nor* sufficient criterion. Not only is there more to translation than language competence, thinking about bilingual competence in terms of some 'word perfect' performance across languages is based on misconceptions of what is involved.

References to 'professional translators' are equally problematic. Arguably this could refer to anyone who earns their living by translating. However, it is often used or taken to imply skills and experience better associated with expert translators, that is, skilled practitioners. Gile (1995:22-23) gives a definition along 'skilled practitioner' lines (not in connection with surveys). However, this in turn raises the issue of what the yardsticks for translator skills can be (cf. Wilss, 1996:147f.). Essentially, translators should have translating skills and translating experience. However, even translation studies literature debates at length what these involve. And while experience helps develop skills, it is no guarantee for them. We consider translating skills to be more important than survey

translation experience, given that guidelines and examples could be provided for translators (but see below).

In survey literature, as in translation studies, views differ on what translators need to know about a topic in order to be able to translate well. It seems unreasonable to require that translators of philosophy must be philosophers and translators of books on calligraphy, calligraphers. On the other hand, in order to choose well between possible translation options, translators need not only to be proficient in the languages but also proficient in ‘reading’ the text and the text type. In other words, translators need to understand the material in order to make informed decisions. Survey translators, therefore, need, for example, a basic understanding of the measurement functions of questionnaires to be able to recognise certain problems (Hambleton, 1993; Hulin, 1987; Borg, this volume) – for which translation will or will not offer a solution. From this follows that sufficient and suitable materials should be provided and explained, so as to help translators produce a satisfactory product (McKay et al., 1996). In the field of survey research, little has been done to develop training or informational materials.

For survey translation, especially perhaps in the multi-lingual context, it is currently unrealistic to expect to find translators who have experience in survey translation, a good understanding of the relevant survey practices and are also in command of both translator skills and proficiency in the languages needed. Within translation studies, opinions differ on how best to go about training translators or assessing their work; nevertheless, a number of basic principles are generally accepted. These could be adapted for survey translation and assessment. Training and informational materials can readily be developed from survey work already done, and new (and old) source questionnaires could be annotated without undue difficulty. The modest annotations in ISSP modules, for example, could be developed systematically, as could a framework for annotating translations for posterity (and for secondary analysis).

7. Conclusion

The goals of questionnaire translation are at present under-defined. The criteria of assessment also remain unarticulated and are, it must be assumed, established on the basis of individual perceptions of 'common sense'. Thus undertaking survey translation may well seem more like setting off on an adventure with unforeseen consequences than anything resembling a systematically organised undertaking.

Questionnaires, on the other hand, *look* easy to translate. After all, questionnaire design handbooks recommend that vocabulary and syntax are kept fairly simple, sentence length is also often short, and the item content of many general population surveys refers to well-known, almost everyday, issues, institutions and entities. In certain senses, questionnaires *are* simple texts. In other respects, some of which have been mentioned here, survey translation is fairly complex. The brevity of items and the quick changes between topics across items mean that preceding sections can rarely be utilised to interpret later sections (cf. different comments on brevity in Sechrest et al., 1972; Hayashi, Suzuki and Sasaki, 1992). Cognitive research has convincingly demonstrated, on the other hand, how respondents extend common 'reading strategies' to questionnaires and thus make links between items not intended by researchers (e.g., Schwarz, 1996). In any case, whether survey translation is relatively simple or not, it involves decisions and selection, and it involves difference as well as equivalence.

When discussed, the process of survey translation is talked about in terms of finding appropriate words, phrases, or sentences in a target language, and about handling grammatical and syntactical features of sentences across languages but rarely in terms of conveying communicative functions of a source text – or source text units – in a target text. As suggested earlier, a focus on communicative function is unlikely to be compatible with literal or close translation as implemented in surveys. It is, on the other hand, central to conveying intended meanings. Whether conveying the intended meaning of a source text item results in a target language question which also taps the intended dimension or construct is a separate issue. At the same time, equivalence of dimensions or constructs to

be measured is *the* essential prerequisite in comparative cross-cultural research. Translators (and secondary analysts), therefore, need information on the dimension/construct supposed to be tapped, as well as an indication of the intended salient reading of the text for each item.

Important challenges to be met in questionnaire translation are similar to those faced in formulating monolingual questionnaires. Cognitive survey research has shown how important both the wording and arrangement of questions (item and response options) and instructions are. Designers formulate, pre-test and re-formulate in order to arrive at the most appropriate expression and arrangements for a given audience and study purpose. Optimal expression of items, instructions, and response scales is one of the tasks also faced by translators, in most cases with considerably less information about the communicative intention than in the monolingual context. Since all questions – not just poorly written ones – are open to different readings, this lack of information compromises translators' decisions about which meaning is *salient* and how best to formulate this in a second language. Without advance task specifications, translators are implicitly setting their own specifications. Providing information and documentation on all these aspects is not standard practice in survey research. Some of it would not be difficult to provide. For other information, the cross-cultural research to match available monocultural research is only beginning. Much remains to be done. Almost thirty years ago, Werner and Campbell (1970) offered to set up a clearing house on translation issues, so as to gather information needed by the scientific community. The need to investigate, document, systematise, accumulate and disseminate information is no less acute today, even if modern technology offers us tools for the job. Without this information, it is difficult to see how the high standards demanded of monocultural item formulation can be extended to decisions about and for translation, or, indeed, *against* 'mere' translation.

References

- Acquadro, C., Jambon, B., Ellis, D. and Marquis, P. (1996). Language and Translation Issues. In: B. Spilker (ed.), *Quality of Life and Pharmacoeconomics in Clinical Trials* (2nd edition). Philadelphia: Lippincott-Raven.
- Alwin, D.F., Braun, M., Harkness, J.A. and Scott, J. (1994). Measurement in Multi-National Surveys. In: I. Borg and P.Ph. Mohler (eds.), *Trends and Perspectives in Empirical Social Research* (pp. 26-39). Berlin: de Gruyter.
- Baker, M. (1992). *In Other Words: A Coursebook on Translation*. London: Routledge.
- Brislin, R.W. (1970). Back-Translation for Cross-Cultural Research. *Journal of Cross-Cultural Psychology* 1: 185-216.
- Brislin, R.W. (1976). Introduction. In: R.W. Brislin (ed.), *Translation: Applications and Research*. (pp. 1-43). New York: Gardner.
- Brislin, R.W. (1980). Translation and Content Analysis of Oral and Written Material. In: H.C. Triandis and J.W. Berry (eds.), *Handbook of Cross-Cultural Psychology*, (vol. 2, pp. 389-444). Boston: Allyn & Bacon.
- Brislin, R.W. (1986). The Wording of Translation of Research Instruments. In: W.J. Lonner and J.W. Berry (eds.), *Field Methods in Cross-Cultural Research* (pp. 137-164). Beverly Hills: Sage.
- Davis, J.A. (1993). *Memorandum to the ISSP*. Chicago: National Opinion Research Center (mimeo).
- Flaherty, J., Moises G.F., Pathak, D., Mitchell, T., Wintrob, R., Richman, J.A. and Birz, S. (1988). Developing Instruments for Cross-Cultural Psychiatric Research. *The Journal of Nervous and Mental Disease* 176(5): 257-263.
- Gile, D. (1995). *Basic Concepts and Models for Interpreter and Translator Training*. Amsterdam: John Benjamins.
- Guillemin, F., Bombardier, C. and Beaton, D. (1993). Cross-Cultural Adaptation of Health-Related Quality of Life Measures: Literature review and proposed guidelines. *Journal of Clinical Epidemiology* 46(12): 1417-1432.
- Gutknecht, C. and Rölle, L. (1996). *Translating by Factors*. Albany: State University of New York Press.
- Gutt, A. (1991). *Translation and Relevance*. Oxford: Basil Blackwell.
- Hambleton, R.K. (1993). Translating Achievement Tests for Use in Cross-National Studies. *European Journal of Psychological Assessment* 9(1): 57-68.

- Hambleton, R.K. (1994). Guidelines for Adapting Educational and Psychological Tests: A progress report. *European Journal of Psychological Assessment (Bulletin of the International Test Commission)* 10: 229-244.
- Harkness, J.A. (1994). Who is Who in Questionnaires? Paper presented at the World Congress of Sociology, Bielefeld.
- Harkness, J.A. (1995). Getting a Word in Edgeways. Questionnaires as Texts and Discourse. Paper presented at the annual meeting of the American Association for Public Opinion Research, Fort Lauderdale, Florida.
- Harkness, J.A. (1995b)
- Harkness, J.A. (1996a). Minding one's P's and Q's and one's P's and C's: Gender issues in questionnaire translation. Paper presented at the World Association for Public Opinion Research meeting, Salt Lake City, Utah.
- Harkness, J.A. (1996b). The (Re)Presentation of Self in Everyday Questionnaires. Paper presented at the International Sociological Association Conference on Social Science Methodology, Colchester.
- Harkness, J.A. (1996c). Thinking Aloud about Survey Translation. Paper presented at the International Sociological Association Conference on Social Science Methodology, Colchester.
- Harkness, J.A. and Braun, M. (in preparation). Text-based and Data-based Perspectives on Questionnaire Translation.
- Harkness, J.A., Mohler, P.Ph. and McCabe, B. (1997). Towards a Manual of European Background Variables. ZUMA Report on Background Variables in a Comparative Perspective. In: J.A. Harkness, P.Ph. Mohler, and R. Thomas, *General Report on Study Programme for Quantitative Research (SPQR)*. Report to the European Commission. Mannheim: ZUMA (mimeo).
- Harkness, J.A., Mohler, P.Ph., Smith, T.W. and Davis, J.A. (1997). *Final Report on the Project: Research into Methodology of Intercultural Surveys (MINTS)*. Mannheim: ZUMA (mimeo).
- Hayashi, C., Suzuki, T. and Sasaki, M. (1992). *Data Analysis for Comparative Social Research: International Perspectives*. Amsterdam: North-Holland.
- Holz-Mänttari, J. (1984). Sichtbarmachung und Beurteilung translatorischer Leistungen bei der Ausbildung von Berufstranslatoren [The elucidation and evaluation of translation performances in translator training]. In: W. Wilss and G. Thome (eds.), *Die Theorie des Übersetzens und ihr Aufschlußwert für die Übersetzungs- und Dolmetschdidaktik* [Translation theory and its relevance for the teaching of translation and interpretation]. Tübingen: Narr.
- Hönig, H. and Kussmaul, P. (1984). *Strategie der Übersetzung*. Tübingen: Narr.

- House, J. (1977). *A Model for Translation Quality Assessment*. Tübingen: Narr.
- Houtkoop-Steenstra, H. (1995). Conversational Problems in Telephone Survey Interviews. Paper presented at the International Conference on Survey Measurement and Process Quality, Bristol.
- Hui, C.H. and Triandis, H.C. (1985). Measurement in Cross-Cultural Psychology. A Review and Comparison of Strategies. *Journal of Cross-Cultural Psychology* 16(2): 131-152.
- Hulin, C.L., Drasgow, F. and Komocar, J. (1982). Applications of Item Response Theory to Analysis of Attitude Scale Translations. *Journal of Applied Psychology* 67(6): 818-825.
- Hulin, C.L. (1987). A Psychometric Theory of Evaluations of Item and Scale Translations: Fidelity across languages. *Journal of Cross-Cultural Psychology* 18(2): 115-142.
- Johnson, T.P., O'Rourke, D., Chavez, N. et al. (1997). Social Cognition and Response to Survey Questions Among Culturally Diverse Populations. In: Lyberg, L., Biemer, P., Collins, M. et al. (eds.), *Survey Measurement and Process Quality* (pp. 87-113). New York: Wiley & Sons
- Kiraly, D. (1995). *Pathways to Translation*. Ohio: Kent University Press.
- Kussmaul, P. (1986). Übersetzung als Entscheidungsprozeß. Die Rolle der Fehleranalyse in der Übersetzungsdidaktik. In: M. Snell-Hornby (ed.), *Übersetzungswissenschaft - Eine Neuorientierung* (pp. 206-229). Tübingen: Francke.
- Kussmaul, P. (1995). *Training the Translator*. Amsterdam: John Benjamins.
- Lyons, J. (1977). *Semantics*, Vol. 1. Cambridge: Cambridge University Press.
- Maier, M.H. (1991). *The Data Game: Controversies in Social Science Statistics*. Armonk, N.Y.: Sharpe.
- McKay, R.B., Breslow, M.J., Sangster, R.L., Gabbard, S.M., Reynolds, R.W., Nakamoto, J.M. and Tarnai, J. (1996). Translating Survey Questionnaires: Lessons Learned. *New Directions for Evaluation* 70: 93-105.
- Newmark, P. (1988). *A Textbook of Translation*. London: Prentice Hall.
- Prieto, A. (1992). A Method for Translation of Instruments to Other Languages. *Adult Education Quarterly* 43(1): 1-14.
- Przeworski, A. and Teune, H. (1970). *The Logic of Comparative Social Inquiry*. New York: John Wiley & Sons.
- Reiss, K. and Vermeer, H.J. (1984). *Grundlegung einer allgemeinen Übersetzungstheorie*. Tübingen: Niemeyer.

- Schaeffer, N.C. (1995). Negotiating Uncertainty: Uncertainty proposals and their disposal in standardised interviews. Paper presented at the International Conference on Survey Measurement and Process Quality, Bristol.
- Schoua, A.S. (1985). *An English/Spanish Test of Decentering for the Translation of Questionnaires*. Unpublished dissertation. Northwestern University.
- Schoua-Glusberg, A.S. (1988). A Focus-Group Approach to Translating Questionnaire Items. Paper presented at the annual meeting of the American Association for Public Opinion Research, Toronto.
- Schoua-Glusberg, A.S. (1989a). The Spanish Version of the NHBQ Questionnaire: Translation Issues. Paper commissioned under Order 263-MD-835040 for presentation at the National Institute of Health AIDS and Sexual Behavior Change Conference.
- Schoua-Glusberg, A.S. (1989b). Using Ethnoscience and Focus Group Techniques in Translation. Paper presented at the annual meeting of the Central States Anthropological Society, South Bend, IN.
- Schoua-Glusberg, A.S. (1992). *Report on the Translation of the Questionnaire for the National Treatment Improvement Evaluation Study*. Chicago: National Opinion Research Center (mimeo).
- Schwarz, N. (1996). *Cognition and Communication. Judgmental Biases, Research Methods, and the Logic of Conversation*. Mahwah: Lawrence Erlbaum.
- Sinaiko, H.W. and Brislin, R.W. (1973). Evaluating Language Translations: Experiments on three assessment methods. *Journal of Applied Psychology* 57(3): 328-334.
- Snell-Hornby, M. (1988). *Translation Studies. An Integrated Approach*. Amsterdam: John Benjamins.
- Sechrest, L., Fay, T.L. and Hafeez Zaidi, S.M. (1972). Problems of Translation in Cross-Cultural Research. *Journal of Cross-Cultural Psychology* 3(1): 41-56. Reprint in: L.A. Samovar and R.E. Porter (eds.) (1988), *Intercultural Communication: A Reader* (5th edition). Belmont: Wadsworth.
- Smith, T.W. (1995). Little Things Matter: A sampler of how differences in questionnaire format can affect survey responses. Paper presented at the annual meeting of the American Association for Public Opinion Research, Fort Lauderdale, Florida.
- Stanley, J.S. (1995a). *Results of Behavior Coding of December Agricultural Survey Interviews. Introduction, Acres Operated, and Hogs and Pigs Sections*. DCB Staff Report Number DCB-95-01. Washington, DC: United States Department of Agriculture. National Agricultural Statistics Service.

- Stanley, J.S. (1995b). *Results of Behavior Coding of December Agricultural Survey Interviews. Corps Section*. DCB Staff Report Number DCB-95-02. Washington, DC: United States Department of Agriculture. National Agricultural Statistics Service.
- Tanzer, N.K., Ellis, B.B., Zhang, H., Sim, C.Q.E., Broer, M. and Gittler, G. (1997). Cross-Cultural Decentering of Test Instructions in a Letter-Cancellation Test: A field test of the ITC Guidelines for test adaptations. In: N.K. Tanzer (Chair): *Advances in Test Translation Methodology*. Symposium conducted at the 5th European Congress of Psychology, Dublin.
- Triandis, H.C. (1994): *Culture and Social Behavior*. New York: McGraw-Hill.
- Van de Vijver, F.J.R. and Hambleton, R.K. (1996). Translating Tests: Some practical guidelines. *European Psychologist* 1(2): 89-99.
- Van de Vijver, F.J.R. and Leung, K. (1997). Methods and Data Analysis of Comparative Research. In: W. Berry, Y.H. Poortinga and J. Pandey: (eds.), *Handbook of Cross-Cultural Psychology* (2nd edition, vol. 1, pp. 257-300). Boston: Allyn & Bacon.
- Vinay, J.P. and Darbelnet, J. (1958, translated 1995). *Comparative Stylistics of French and English. A Methodology for Translation*. Amsterdam: John Benjamins.
- Werner, O. and Campbell, D. (1970). Translating, Working through Interpreters and the Problem of Decentering. In: R. Naroll and R. Cohen (eds.), *Handbook of Cultural Anthropology*. New York: American Museum of Natural History.
- Wilss, W. (1996). *Knowledge and Skills in Translator Behavior*. Amsterdam: John Benjamins.